INFORMATION AND CONTROL 10, 394–408 (1967)

Detection of an Unknown Waveform Randomly Recurring in Gaussian Noise*†

MELVIN HINICH

Carnegie Institute of Technology, Pittsburgh, Pennsylvania, 15213 and Hudson Laboratories of Columbia University, Dobbs Ferry, New York 10522

Matched filtering is a widely used signal processing technique for detecting a known signal imbedded in noise. In the pulsed sonar case, the medium often distorts the received signal to such an extent that matched filtering is not optimal. This paper deals with the case where a record is obtained of Gaussian noise containing many repetitions of the distorted signal waveform. After obtaining an estimate of the waveform autocorrelation based upon nonoverlapping sections from the first part of the record, a test procedure is developed for determining whether or not the waveform is present in a subsequent section. The test procedure is derived from a Taylor series approximation of the likelihood ratio, where the signal-to-noise ratio is quite low and the probability that two or more waveforms are in the section is negligible. This suboptimal test involves cross correlating the sample autocorrelation with the waveform autocorrelation.

An approximation is given to the difference in power between the likelihood-ratio test and its approximation. A comparison is made with the matched filter approach.

1. INTRODUCTION

Matched filtering is a widely used signal processing technique for detecting the arrival of a known sonar signal imbedded in ambient sea noise. However, due to the instability and nonisotropic nature of the medium it often happens that the sonar waveform is so distorted that the matched filtering is no longer optimal. If the medium remains stable

* Hudson Laboratories of Columbia University Contribution No. 278.

† This work was begun at Carnegie Institute of Technology under support of the U. S. Office of Naval Research, Nonr 760 (24), and completed at the Hudson Laboratories under Nonr 266 (84). Reproduction in whole or in part is permitted for any purpose of the United States Government.

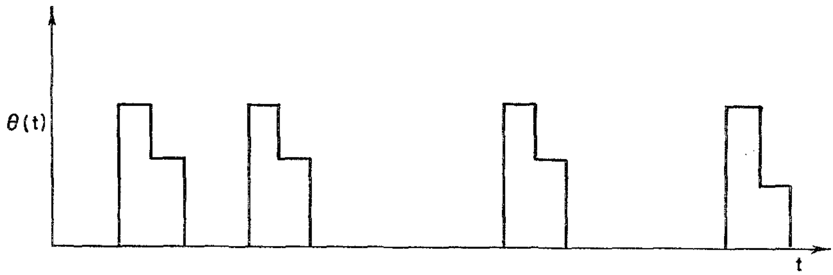

FIG. 1. Recurring waveform

long enough to send and receive a reasonable number of signals, it may be possible to utilize some sort of adaptive waveform detector (Williams, 1966).

Suppose we have an input $X(t)$ made up of an unknown waveform $\theta(t)$ of known length, which is repeated randomly and is imbedded in Gaussian noise with a known covariance function (Fig. 1). The average time between recurrences of the waveform is large compared to its length. In addition, the signal-to-noise ratio is quite low. We wish to determine the times of occurrence of the waveform in a record of discrete-time observations of $X(t)$.

In a recent paper, Jakowatz and Stutt (1960) discuss the present state of development of a special system, the Adaptive Filter (Hinich, 1962; Jakowatz *et al.*, 1960) which utilizes an iterative scheme based upon a cross-correlation type decision rule to detect the times of recurrence of an unknown $\theta(t)$. In related work, Scudder (1965) makes ingenious use of a Bayesian approach to develop another type of iterative scheme that asymptotically behaves as a matched filter. An application of adaptive filtering for sonar is given in (Williams, 1966).

If the noise power is large relative to the waveform, these adaptive schemes fail to converge to a matched filter because of the relatively large false-alarm probability associated with a decision that the waveform $\theta(t)$ is present in a section of the record along with the noise, when in fact it is not true. Besides triggering falsely on just noise alone in the section, the cross-correlation detector can be fooled by a sample correlation function exceeding the threshold of the decision rule because of the situation where a part of $\theta(t)$—the lead end or the tail end, for example—is present in the section along with the noise. If the relative maxima of the waveform autocorrelation function $\psi(\tau) = \int \theta(t + \tau)\theta(t)\, dt$ are not much smaller than the absolute maximum $\psi(0)$, the filter will have difficulty in precisely estimating the times of recurrence of $\theta(t)$. The iterative procedure uses sections of the record where the waveform is present along with the noise in order to update the running estimate of θ, which is then utilized in the decision rule that detects later recurrences of the waveform. Therefore, averaging sections where there is great ambiguity in waveform position causes "blurring" of the estimate of $\theta(t)$ and consequently a severe weakening of the decision procedure. Averaging sections of noise alone does not cause this type of problem since, because of stationarity in the mean, the noise "washes out" in the averaging. However, if the signal power is low, then the time-position ambiguity can foul the iteration and result in instability of the procedure. Moreover, in this case it is difficult to get the initial crude estimate of $\theta(t)$ which primes the filter.

Due to the assumed conditions, it is inappropriate to detect the occurrences of the waveform by simple cross correlation. This paper deals with the extreme case of low "generalized signal-to-noise ratio" (see Section 2) and uses a large-sample approach in order to help obtain general bounds of the asymptotic performance of adaptive waveform detectors. It also suggests an alternative to simple cross correlation for the data processing of the observed $X(t)$ to be used in the decision part of the "detect-estimate-detect" iteration.

Suppose we restrict ourselves to a sampling scheme by which we observe a number of nonoverlapping sections of the record, which are spaced in time sufficiently far apart so that the observations in one section are independent of those in other sections (Fig. 2).

There are two basic possibilities for a given section of $X(t)$ [discussed in detail by Hinich (1965)]. Either that section consists of just noise alone, or there is some part of θ in the section (the head end, tail end, or all of θ). Since the times of occurrence of θ are assumed to have a purely random distribution with an average time between θ's which is longer than the duration of the waveform and the length of the section, then all possible shifted positions of θ in the section are equally likely. Moreover many sections will consist of just noise.

Let the column vector X denote a discrete-time section of $X(t)$ consisting of w successive observations $X_i = X(t_i)$ for $i = 1, \cdots, w$ where $W = t_{i+1} - t_i$ is the sampling interval. By using the prime to denote the transpose, $X' = (X_1, \cdots, X_w)$ is the section, $N' = (N_1, \cdots, N_w)$ is the noise vector, and $\theta' = (\theta_1, \cdots, \theta_n)$ represents the discrete unknown waveform.

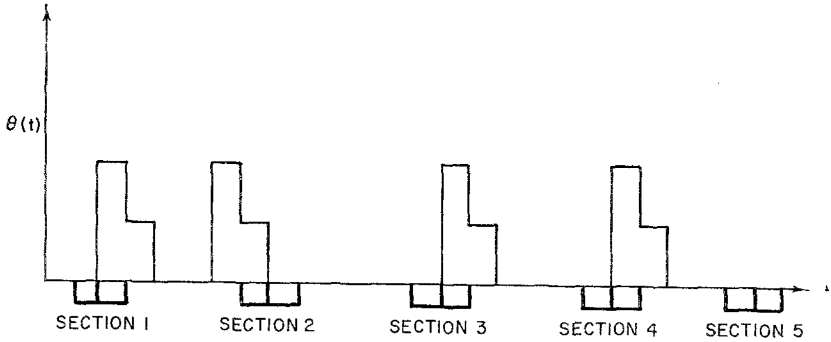

FIG. 2. Sampling scheme—noise removed

It is easy to check that there are $n + w - 1$ possible shifted positions of θ in X. The jth shift can be represented as

$$(S_j\theta)' = (\theta_{j+1}, \cdots, \theta_{j+w})$$

with $\theta_k = 0$ if $k \leqq 0$ or $k \geqq n + 1$. For example $(S_{n-1}\theta)' = (\theta_n, 0, \cdots, 0)$ and $(S_{-w+1}\theta)' = (0, 0, \cdots, \theta_1)$. There is no way of knowing in advance which case is occurring. Since the signal power is low we will not attempt to determine which case, if any, occurred in a given section. Instead we will try to distinguish between the two basic possibilities,

$$\begin{aligned} &H_0 : X = N \\ &H_1 : X = N + S_j\theta \quad \text{for some} \quad j = w + 1, \cdots, n - 1 \end{aligned} \tag{1}$$

i.e., H_0 is the noise alone case and H_1 is the case where some part of θ is in the section.

The noise vector has a w-dimensional multivariate normal distribution with mean zero and a known covariance matrix Σ. We express this by $\mathcal{L}\{N\} = \mathfrak{N}(0, \Sigma)$ where $\mathcal{L}\{N\}$ is the distribution function of the random variable N (Wilks, 1962, Chapt. 7).

Given a section X, we will decide between H_0 and H_1 on the basis of a test derived using a Taylor series approximation of the appropriate likelihood-ratio. Middleton (1960, Chapt. 20) calls such an approximate likelihood-ratio test, *suboptimal*. The difference in power between the likelihood-ratio test and the suboptimal test, will be approximated.

2. ESTIMATION OF THE WAVEFORM AUTOCORRELATION

In the discrete case, the waveform autocorrelation is defined as the

n-dimensional vector $\psi' = (\psi_1, \cdots, \psi_n)$, where

$$\psi_1 = \tfrac{1}{2}\sum_1^n \theta_k^2, \qquad \psi_2 = \sum_1^{n-1} \theta_k\theta_{k+1}, \quad \cdots, \psi_n = \theta_1\theta_n. \tag{2}$$

In order to simplify the testing problem we assume that ψ_0, the DC value of θ, is $\psi_0 = \sum_1^n \theta_k = 0$. This is a reasonable assumption in most applications.

Suppose that we have many non-overlapping sections of $X(t)$ such that the possibilities are given as in (1) for each section. We find it convenient to make the linear transformation on a section X,

$$Z = \Sigma^{-1}X, \tag{3}$$

where Σ^{-1} is the inverse of Σ, the covariance matrix of the Gaussian noise vector N. From (1) we have the two basic possibilities

$$\begin{aligned} &H_0 : Z = N^* \\ &H_1 : Z = N^* + \Sigma^{-1}S_j\theta \quad \text{for some} \quad j = w + 1, \cdots, n - 1, \end{aligned} \tag{4}$$

where $\mathcal{L}\{N^*\} = \mathfrak{N}(0, \Sigma^{-1})$.

Define $Y(z)' = (Y_1(z), \cdots, Y_n(z))$ by

$$Y_k(z) = \sum_{i=1}^{w-k+1} (z_i z_{i+k-1} - \sigma^{i,i+k-1}) \tag{5}$$

for $k = 1, \cdots, n$ with notation $\Sigma^{-1} = (\sigma^{ij})$.

The vector $Y(z)$ is a slightly different form for the discrete *sample autocorrelation* of the transformed vector of observations Z.

It is understood that $z_i = 0$ if $i \leqq 0$ or $i \geqq w + 1$, and thus if $w < n$

$$Y_{w+1}(z) = \cdots = Y_n(z) = 0.$$

It was shown (Hinich, 1965) that for small signal power, it is very difficult to estimate θ while it is relatively easy to estimate ψ from the sections. Furthermore the ψ_i are natural parameters in the approximation to the likelihood-ratio test.

Let $X^{(1)}, \cdots, X^{(m)}$ denote m sections of $X(t)$, γ denote the probability that some part of θ is present in a section, D denote the covariance matrix of the sample autocorrelation $Y(Z)$ given that Z is just noise alone, and $Z^{(k)} = \Sigma^{-1}X^{(k)}$ for every k. Then, adapting the proof of Theorem 4 of Hinich (1965),

$$\hat{\psi} = \frac{n + w - 1}{m\gamma} D^{-1} \sum_{k=1}^{m} Y(Z^{(k)}) \tag{6}$$

is asymptotically (for large m) normally distributed with mean ψ and with a covariance matrix which is as small as possible—to roughly summarize the concept of asymptotic efficiency which is discussed in detail in Chapt. 12 of Wilks (1962). The covariance matrix of $\hat{\psi}$ is of the order of m^{-1} and we suppose that we can observe $X(t)$ sufficiently long so that we can assume the ψ_i as known parameters in the problem of testing for the presence of θ in a section X.

The signal-to-noise ratio was defined (Hinich, 1965) as $R_\theta = \| \theta \|^2/n\sigma^2$ where $\sigma^2 = EN^2(t)$ is the variance of the noise $N(t)$, and $\| \theta \|^2 = \sum_i \theta_i^2$. R_θ was assumed to be small. It is more meaningful in this type of problem to define the signal-to-noise ratio in terms of the noise covariance Σ. Since Σ is positive definite, it has a nonsingular symmetric square root $\Sigma^{1/2}$ such that $(\Sigma^{1/2})^2 = \Sigma$. Define the vector $X^* = (\Sigma^{1/2})^{-1}X$, which is the "prewhitened" version of X since from (1) the noise component of X^* has a diagonal covariance matrix with diagonal elements all equal to σ^2. The maximum power of the waveform component of X^* is $\| \Sigma^{-1/2}\theta \|^2 = \theta'\Sigma^{-1}\theta$.

Let λ_0 be the minimum eigenvalue of Σ, i.e., the maximum eigenvalue of Σ^{-1}. Then the maximum waveform component in X^* is bounded as follows:

$$\theta'\Sigma^{-1}\theta \leqq \lambda_0^{-1}\| \theta \|^2 = \lambda_0^{-1} \sum_i \theta_i^2.$$

This motivates the following generalization for the definition of the signal-to-noise ratio $R_{\theta,\Sigma}$ of the waveform in noise:

$$R_{\theta,\Sigma} = \| \theta \|^2/n\lambda_0 . \tag{7}$$

For example if Σ was almost singular, i.e., λ_0 is near zero, then λ_0^{-1} is large and thus $R_{\theta,\Sigma}$ can be large even though $\| \theta \|$ is small in magnitude. Heuristically, the smallness of λ_0 makes the noise process $N(t)$ very predictable and thus even a waveform with small power can be detected with probability near one.

If the noise is white, the generalized definition reduces to R_θ since Σ is a diagonal matrix with diagonal elements σ^2.

3. STATISTICAL DETECTION OF THE WAVEFORM

We will discuss the likelihood-ratio test of the following two hypotheses for the observed random vector $Z = \Sigma^{-1}X$.

$$\begin{aligned} H_0 &: Z = N^* \\ H_1 &: Z = N^* + \Sigma^{-1}S_j\theta \quad \text{for some} \quad j = -w + 1, \cdots, n - 1, \end{aligned} \tag{8}$$

where $\mathfrak{L}\{N^*\} = \mathfrak{N}(0, \Sigma^{-1})$. In other words, we wish to determine whether or not the section X contains part of the waveform along with noise.

Let $f(z \mid \theta)$ be the density of Z parameterized by θ, given that hypothesis H_1 is true. Since we are averaging over the position uncertainty,

$$f(z \mid \theta) = \frac{1}{n + w + 1} \sum_{j=-w+1}^{n-1} n(z \mid \Sigma^{-1} S_j \theta, \Sigma^{-1}), \tag{9}$$

where $n(z \mid \Sigma^{-1} S_j \theta, \Sigma^{-1})$ is a w-dimensional normal density function with mean $\Sigma^{-1} S_j \theta$ and covariance matrix Σ^{-1}.

The density function of Z given that H_0 is true, is simply $f(z \mid 0) = n(z \mid 0, \Sigma^{-1})$.

Adapting Lemma 1 from a previous paper (Hinich, 1965) and since $\psi_0 = \Sigma\theta_i = 0$, we have the likelihood ratio

$$\begin{aligned} \frac{f(z \mid \theta)}{f(z \mid 0)} = 1 + \frac{1}{n + w - 1} \sum_{k=1}^{n} Y_k(z)\psi_k \\ + \sum_{i,j,k}^{w} G_3(z \mid i, j, k) O(\| \theta \|^3) + K(z, \theta) O(\| \theta \|^4), \end{aligned} \tag{10}$$

where

$$\begin{aligned} G_3(z \mid i, j, k) &= z_i z_j z_k - \sigma^{ij} z_k - \sigma^{ik} z_j - \sigma^{jk} z_i \\ E_0\{G_3(Z \mid i, j, k)\} &= 0 \quad \text{for all} \quad i, j, k \end{aligned} \tag{11}$$

and $E_0\{K^r(Z, 0)\}$ exists and is bounded by some number independent of θ for each $r \geqq 0$. E_0 represents expectation with respect to $f(z \mid 0)$. Moreover the $O(\| \theta \|^3)$ and $O(\| \theta \|^4)$ terms are functions of θ which do not involve z or w.

We observe from (10) that the linear term is zero in this approximation of the likelihood ratio and that the quadratic term is an inner product of the sample correlations $Y(Z)$ and the waveform autocorrelation ψ, which is assumed to be known.

In order to obtain the distribution of the test statistic given below, we will restrict the noise $N(t)$ to be Gaussian Markov with variance normalized to one, i.e., assume that the covariance function of $N(t)$ is

$$EN(t + \tau)N(t) = \rho(\tau) \qquad 0 < \rho < 1 \tag{12}$$

and therefore the noise covariance matrix is

$$\Sigma = \begin{pmatrix} 1 & \lambda & \lambda^2 & \cdots & \lambda^{w-1} \\ \lambda & 1 & \lambda & \cdots & \lambda^{w-2} \\ \lambda^2 & \lambda & 1 & \cdots & \lambda^{w-3} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \lambda^{w-1} & \lambda^{w-2} & \lambda^{w-3} & \cdots & 1 \end{pmatrix},$$

where $\lambda = \rho^W$ and W is the fixed time between successive observations $X(t_i)$ and $X(t_{i+1})$, i.e., W^{-1} is the sampling rate.

From a result by Dixon (1944), we have for large w that the minimum eigenvalue of Σ is $1 - |\lambda|$ and thus from (7) the signal-to-noise ratio is

$$R_{\theta.\Sigma} = \|\theta\|^2/n(1 - |\lambda|).$$

We now prove the following theorem.

THEOREM 1. *For small $R_{\theta,\Sigma}$ the suboptimal likelihood ratio test for choosing between the hypotheses*

$$H_0 : Z = N^* \quad (\textit{noise alone in record})$$

$$H_1 : Z = N^* + \Sigma^{-1} S_j \theta \quad \textit{for some} \quad j = -w + 1, \cdots, n - 1$$

is to reject H_0 if the statistic

$$T(Z) = \frac{1}{n + w - 1} \sum_1^n Y_k(z)\psi_k \tag{13}$$

is greater than a threshold t_α, defined such that

$$\Pr\{T(Z) > t_\alpha \mid H_0\} = \alpha$$

is the false-alarm probability.

For large w the test statistic is approximately normally distributed as follows:

(i) Given that H_0 is true

$$\mathcal{L}\{wT(Z)\} \doteq \mathfrak{N}(0, \sigma_0^2)$$
$$\sigma_0^2 = \psi' D \psi,$$

where

$$D = E_0 Y(Z) Y(Z)'$$

is the $n \times n$ covariance matrix of $Y(Z)$, as is given in Example C of a previous paper (Hinich, 1965).

(ii) Given that H_1 is true

$$\mathcal{L}\{wT(Z)\} \doteq \mathfrak{N}(\mu, \sigma_1^2)$$

$$\mu = \frac{1}{n + w - 1} \psi' D\psi$$

$$\sigma_1^2 = \psi' D\psi + O(\| \theta \|^6),$$

where the $O(\| \theta \|^6)$ term is of order w.

Proof. We obtain the statistic $T(Z)$ from the likelihood ratio (10) by neglecting the $O(\| \theta \|^3)$ and higher terms.

From Example C (Hinich, 1965) it follows that for large w the components of the matrix $w^{-1}D$ are $O(1)$. Furthermore, $w \doteq n + w - 1$. Then by adapting Lemmas 5 and 6 of Hinich (1965), we have

$$E_\theta\, Y(Z) = \frac{1}{n + w - 1} D\psi \tag{14}$$

and

$$\operatorname{Cov}_\theta Y(Z) = D + O(\| \theta \|^2), \tag{15}$$

where the term $O(\| \theta \|^2)$ is of order w.

Since $(n + w - 1)T(Z) = Y(Z)'\psi$, the expressions for μ and σ_1^2 follow from (14) and (15). Under the null hypothesis H_0,

$$E_0 Y(Z) = 0$$

$$\operatorname{Cov}_0 Y(Z) = D$$

and thus σ_0^2 follows.

The asymptotic normality of the statistic follows from the main theorem of Anderson and Walker (1964).

The $n + w - 1$ is used instead of w in the expression for the mean μ in order to preserve the "small sample" unbiasedness of the statistic $T(E_\theta T(Z) = \mu)$. Of course for large w, $n + w - 1 \doteq w$.

If the noise is white ($\lambda = 0$) we have a somewhat sharper result.

THEOREM 2. *Given that $\Sigma = I$, the identity matrix, for small $\| \theta \|$ the sub-optimal likelihood-ratio test is to reject H_0 if the statistic*

$$T(X) = \frac{1}{n + w - 1} \sum_{k=1}^{n} Y_k(X)\psi_k$$

is greater than t_α, where

$$Y_1(X) = \sum_{i=1}^{w} X_i^2 - w$$

$$Y_2(X) = \sum_{i=1}^{w-1} X_i X_{i+1}$$

$$\vdots \qquad \vdots$$

$$Y_n(X) = \sum_{i=1}^{w-n+1} X_i X_{i+n-1}.$$

For large w the test statistic is approximately normally distributed as follows:

(i) Given that H_0 is true

$$\mathcal{L}\{T(Z)\} \doteq \mathfrak{N}(0, \sigma_0^2)$$

$$\sigma_0^2 = w^{-1}\left(2\psi_1^2 + \sum_{i=2}^{n} \psi_i^2\right).$$

(ii) Given that H_1 is true

$$\mathcal{L}\{T(Z)\} \doteq \mathfrak{N}(\mu, \sigma_1^2)$$

$$\mu = (n + w - 1)^{-1}\left(2\psi_1^2 + \sum_{i=2}^{n} \psi_i^2\right)$$

$$\sigma_1^2 = w^{-1}\left[\left(2\psi_1^2 + \sum_{i=2}^{n} \psi_i^2\right) + O(\| \theta \|^6)\right],$$

where the term $O(\| \theta \|^6)$ is of order one in w.

Proof. Since $\Sigma = I$, $Z = X$. Furthermore

$$D = \begin{pmatrix} 2w & & & & 0 \\ & w-1 & & & \\ & & w-2 & & \\ & & & \ddots & \\ 0 & & & & w-n+1 \end{pmatrix} \qquad (16)$$

Therefore

$$\lim_{w\to\infty} w^{-1}\psi D\psi' = 2\psi_1^2 + \psi_2^2 + \cdots + \psi_n^2.$$

From Theorem 1, for large w the distribution under H_0,

$$\mathcal{L}\{T(X)| H_0\} \doteq \mathfrak{N}(0, w^{-2}\sigma_0^2)$$

and under H_1,

$$\mathcal{L}\{T(X)| H_1\} \doteq \mathfrak{N}(w^{-1}\mu, w^{-2}\sigma_1^2).$$

The desired results then follow where once again we use $n + w - 1$ instead of w in μ in order to keep $E_\theta T(X) = \mu$.

Notice that by neglecting higher order terms, the detectability μ is

$$\mu = \sigma_0^2 = \sigma_1^2 = O(\| \theta \|^4).$$

By assumption this is very small and thus it is very hard to separate signal from noise in one observation section X. We can increase the detectability if we observe a large number of independent sections $X^{(1)}, \cdots, X^{(m)}$ where each section has the two basic possibilities as in (8) but where the reality is unknown in each case. By taking the logarithm of the likelihood and expanding as above, it is easy to see that the suboptimal procedure is to reject H_0 if

$$T(X^{(1)}, \cdots, X^{(m)}) = \frac{1}{m(n + w - 1)} \sum_{i=1}^{m} \sum_{k=1}^{n} Y_k(X^{(i)})\psi_k \qquad (17)$$

is greater than t_α where under H_0

$$\Pr\{T(X^{(1)}, \cdots, X^{(m)}) > t_\alpha \mid H_0\} = \alpha.$$

For sufficiently large m the power of the test (the probability of rejecting H_0 when H_1 is true) can be made as close to one as desired, i.e., $m \rightarrow \infty$,

$$\Pr\{T > t_\alpha \mid H_1\} \rightarrow 1,$$

which implies that the probability of missing the waveform goes to zero.

4. POWER OF THE TEST

In this section we will approximate the difference in power between the likelihood-ratio test and its approximation, assuming white noise.

The likelihood-ratio test with false-alarm probability α is to reject H_0 if

$$\lambda(x \mid \theta) = \frac{f(x \mid \theta)}{f(x \mid 0)} > h_\alpha, \qquad (18)$$

where

$$\Pr\{\lambda(X \mid \theta) > h_\alpha \mid H_0\} = \alpha.$$

From (10) the higher order terms have variance $O(w \| \theta \|^6)$ and thus

$$\lambda(x \mid \theta) = 1 + T(x) + O(w^{-1/2} \| \theta \|^3). \tag{19}$$

Since both tests have false-alarm probability α,

$$h_\alpha = 1 + t_\alpha + O(w^{-1/2} \| \theta \|^3). \tag{20}$$

Let $\beta_\lambda(\theta)$ be the power of the λ test, i.e., it is the probability of deciding that there is some part of θ present in X when in fact it is true. Thus

$$\beta_\lambda(\theta) = \Pr\{\lambda(X \mid \theta) > h_\alpha \mid H_1\}. \tag{21}$$

The optimality of the likelihood-ratio test can be expressed in terms of the power. Given any other test with false-alarm probability α and power β, $\beta \leqq \beta_\lambda$ for each θ.

Let $\beta_T(\theta)$ be the power of the T test, i.e.,

$$\beta_T(\theta) = \Pr\{T(X) > t_\alpha \mid H_1\}. \tag{22}$$

We now prove the following results:

THEOREM 3. *$\beta_T(\theta) \leqq \beta_\lambda(\theta) \leqq \beta_T(\theta) + O(\| \theta \|)$ for small $\| \theta \|$ and large w, with the error term of order one in w.*

Proof. If Y is a normal random variable with mean m and variance v^2, then by taking derivatives

$$\Pr\{Y > t - \epsilon\} = \Pr\{Y > t\} + O\left(\frac{\epsilon}{v}\right) \tag{23}$$

for small ϵ.

Substituting (19) and (20) in (22),

$$\begin{aligned} \beta_\lambda &= \Pr\{\lambda(X \mid \theta) > 1 + t_\alpha + O(w^{-1/2} \| \theta \|^3) \mid H_1\} \\ &= \Pr\{T(X) > t_\alpha + O(w^{-1/2} \| \theta \|^3) \mid H_1\}. \end{aligned} \tag{24}$$

By Theorem 2, $T(X)$ is approximately normal with variance of the order $O(w^{-1} \| \theta \|^4)$. Thus from (23) and (24),

$$\begin{aligned} \beta_\lambda &= \Pr\{T(X) > t_\alpha \mid H_1\} + O(\| \theta \|) \\ &= \beta_T + O(\| \theta \|). \end{aligned}$$

5. MATCHED FILTER COMPARISON

Returning to the detection problem as formulated in the beginning of Section 3, we will compare the suboptimal detector with a version of matched filtering.

Suppose that we are given the waveform vector θ and the information that if θ is present in the section X, it has the kth shifted position $S_k\theta$ for a fixed $-w + 1 \leqq k \leqq n - 1$. Thus the two alternatives for $Z = \Sigma^{-1}X$ are

$$\begin{aligned} H_0 &: Z = N^* \\ H_1' &: Z = N^* + \Sigma^{-1}S_k\theta, \end{aligned} \tag{25}$$

where $\mathcal{L}\{N^*\} = \mathfrak{N}(0, \Sigma^{-1})$. For example if $w = n$ and $k = 0$, the H_1' hypothesis states that $S_0\theta = \theta$ is centered in the section. Given the formulation as in (25), the likelihood-ratio test is to reject H_0 if the statistic $S(Z) = Z'S_k\theta$ is greater than some threshold. This is a form of matched filtering, which can be easily seen if we assume the noise is white, $w = n$, and $k = 0$, for then the test statistic becomes

$$X'\theta = \sum_{i=1}^{n} \theta_i X_i .$$

Given H_0 is true, $S(Z)$ is normally distributed with *mean zero* and variance

$$\sigma^2 = (S_k\theta)'\Sigma^{-1}S_k\theta. \tag{26}$$

Given H_1 is true, $S(Z)$ is also normally distributed with *mean* σ^2 and *variance* σ^2.

However, suppose we were misinformed and that we observe a section Z which is distributed as given in (8), i.e., if the waveform is present in Z then all shifted positions of θ are equally likely. Since we believe that Z is distributed as in (25), we reject H_0 if $S(Z) = Z'S_k\theta$ is greater than some threshold.

Given H_0 is true, $S(Z)$ is normally distributed with *mean zero* and variance σ^2 as in (26). Because Z is distributed as in (8), given H_1, $S(Z)$ is also normal but with mean

$$\mu = \frac{1}{n + w - 1} \sum_{j=-w+1}^{n-1} (S_j\,\theta)'\Sigma^{-1}S_k\,\theta. \tag{27}$$

However, it can be easily shown that

$$\sum_{j=-w+1}^{n-1} (S_j\theta)' = (\psi_0, \psi_0, \cdots, \psi_0); \tag{28}$$

and since $\psi_0 = \Sigma\theta_i = 0$, applying (28) to (27) we have that $\mu = 0$. Therefore even given H_1, $S(Z)$ has zero mean.

The variance of $S(Z)$ given H_1, is

$$\begin{aligned}\sigma^{*2} &= \frac{1}{n+w-1} \sum_{j=-w+1}^{n-1} (S_k\,\theta)'[\Sigma^{-1} + \Sigma^{-1}S_j\,\theta(S_j\,\theta)'\Sigma^{-1}]S_k\,\theta \\ &= (S_k\,\theta)'\Sigma^{-1}S_k\,\theta + \frac{1}{n+w-1} \sum_{j=-w+1}^{n-1} [(S_k\,\theta)'\Sigma^{-1}S_j\,\theta]^2.\end{aligned} \tag{29}$$

Now by the Schwarz inequality

$$\begin{aligned}[(S_k\theta)'\Sigma^{-1}S_j\theta]^2 &\leqq [(S_k\theta)'\Sigma^{-1}S_k\theta][(S_j\theta)'\Sigma^{-1}S_j\theta] \\ &\leqq \lambda_0^{-2}[(S_k\theta)'S_k\theta][(S_j\theta)'S_j\theta],\end{aligned} \tag{30}$$

where λ_0 is the minimum eigenvalue of Σ. It can easily be shown that

$$\sum_{j=w+1}^{n-1} (S_j\theta)'S_j\theta = 2\psi_1 = \|\theta\|^2 \tag{31}$$

and also

$$(S_k\theta)'S_k\theta \leqq \|\theta\|^2.$$

Applying (30) and (31) to (29), we have

$$\sigma^{*2} \leqq \sigma^2 + \lambda_0^{-2}\|\theta\|^4. \tag{32}$$

Summing up the above

$$\begin{aligned}\mathcal{L}\{S(Z)\,|\,H_0\} &= \mathfrak{N}(0, \sigma^2) \\ \mathcal{L}\{S(Z)\,|\,H_1\} &= \mathfrak{N}(0, \sigma^2 + \lambda_0^{-2}\|\theta\|^4).\end{aligned}$$

We observe that the only difference in the distribution of the test statistic is an increase in the variance of order $\|\theta\|^4$ given the alternative H_1. This increase in variance is of the magnitude of the mean shift given H_1 for the suboptimal test procedure. In order to take advantage of the increase in variance of $S(Z)$ given H_1, we would compute $S^2(Z)$ and compare it with a threshold, and this involves the same sort of data processing as in the suboptimal scheme plus knowledge about θ which is difficult to obtain.

RECEIVED: December 8, 1966

REFERENCES

ANDERSON, T. W., AND WALKER, A. M. (1964), On the asymptotic distribution of the autocorrelations of a sample from a linear stochastic process. *Ann. Math. Statist.* **35,** 1296–1303.

DIXON, W. J. (1944), Further contributions to the problem of serial correlation. *Ann. Math. Statist.* **15,** 119–144.

HINICH, M. (1962), A model for a self-adapting filter. *Inform. Control* **5,** 185–203.

HINICH, M. (1965), Large-sample estimation of an unknown discrete waveform which is randomly repeating in Gaussian noise. *Ann. Math. Statist.* **36,** 489–508.

JAKOWATZ, C., SHUEY, R., AND WHITE, G. (1960), Adaptive waveform recognition. "Information Theory—Fourth London Symposium" (Colin Cherry, ed.), pp. 317–326, Butterworths, Washington, D. C.

JAKOWATZ, C., AND STUTT, C. (1964), Adaptive filter. *Underwater Acoustics* (U. S. Navy Journal), **14–4,** 767–775.

MIDDLETON, D. (1960), "Introduction to Statistical Communication Theory." McGraw-Hill, New York.

SCUDDER, H. J. (1965), Adaptive communication receivers. *IEEE Trans. Inform. Theory* **IT 11–2,** 167–174.

WILKS, S. (1962), "*Mathematical Statistics*." Wiley, New York.

WILLIAMS, R. E. Adaptive correlation of time distorted signals. Contribution No. 266, Hudson Laboratories of Columbia University, Dobbs Ferry, N. Y., publication pending.